DOCUMENT RESUME

ED 058 290 TM 001 005

AUTHOR Murray, Norman B.
TITLE Measurement of a Learner's Achievement of a Collection of Behavioral Objectives.
PUB DATE Apr 71
NOTE 11p.; Paper presented at the Annual Meeting of the California Educational Research Association, San Diego, California, April 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement; *Behavioral Objectives; Classification; *Evaluation Criteria; *Evaluation Techniques; Feedback; Grading; Hypothesis Testing; Measurement Techniques; Performance Criteria; Performance Specifications; Research Methodology; *Summative Evaluation; Systems Analysis; *Systems Approach

ABSTRACT

Evaluation of an individual's progress in an academic or training program requires evaluation of his achievement of a collection of behavioral objectives. The nature of the terminal behavior often imposes a hierarchy on the enabling and entering behaviors that can be used to lend additional meaning to classification of the learner's performance, i.e., to the grades assigned during and at the end of the course, in the case of an academic program. The following discussion suggests ways more meaningful evaluation can be accomplished, meaningful in the sense that the resulting classifications, i.e., grades, imply definite degrees of progress up the behavioral hierarchy. (Author)

MEASUREMENT OF A LEARNER'S ACHIEVEMENT OF A COLLECTION OF BEHAVIORAL OBJECTIVES

Norman B. Murray

Graduate Student

Brigham Young University

Provo, Utah

Presented at the Annual Convention

of the California Educational Reseach Association

April 29, 1971

Introductory Statement

Evaluation of an individual's progress in an academic or training program requires evaluation of his achievement of a collection of behavioral objectives. The nature of the terminal behavior often imposes an hierarchy on the enabling and entering behaviors that can be used to lend additional meaning to classification of the learner's performance, i.e., to the grades assigned during and at the end of the course, in the case of an academic program. The following discussion suggests ways more meaningful evaluation can be accomplished, meaningful in the sense that the resulting classifications, i.e., grades, imply definite degrees of progress up the behavioral hierarchy.

Measuring Achievement of One Behavioral Objective

Since the following discussion is based on the notion of measuring the achievement of one behavioral objective we shall say a few words about that first. This task might be visualized as performing a small experiment to find out if someone is able to perform the specified behavior. Such an experiment should be designed using the same procedures as are followed in conducting any empirical investigation.

The first step in any experiment is to define the problem under investigation. In the case of measuring someone's achievement of a behavioral objective the problem might be stated in the form of a question, "Can the given subject perform the specified behavior?" As part of the task of stating the problem, an operational definition of the specified behavior must be provided. In this way we are able to set up an experimental situation in which the behavior can be measured or observed. The operational definition of the behavior would determine the physical situation in which the behavior would occur.

It is not hard at all to set down a definition of an overt behavior like shooting a basketball, but for other behaviors specifying an operational definition becomes more difficult. For example, what would it be if a behavior involved identifying unencountered instances of an impacted tooth? Or, what would it be if the behavior involved use of the classical laws of motion in a practical situation? In any case, an operational definition must be provided in order to have a valid measure of the behavior.

The next step in this small experiment set up to measure the achievement of one behavioral objective is formulating an hypothesis. If the hypothesis states that the subject can perform the given behavior, then under this assumption we are able to proceed in setting up an experiment in which, if the subject fails to perform the behavior, the hypothesis can be rejected.

Before the experimental method is designed and described, a permissible level of significance must be specified and also a permissible level of Type II error just as is done in any experiment based on a statistical method.

The next step in this small experiment is defining and describing the experimental method. The type of subject or the subject should be taken into account. This must be done in order to be able to count on behaviors the subject is already able to perform. The apparatus and materials involved in this experiment should also be described so that someone else wanting to duplicate the experiment could reproduce the materials and build the apparatus.

All of the phases described to this point form a foundation for the heart of the experiment, that is, the experimental procedure. Before the experiment can actually be performed, it must be designed and described in such a way that an independent person could duplicate the experimental procedure. Procedures must be described accurately and in detail. Criteria must be defined under which the experiment would be terminated. In the case of measuring achievement of one behavioral objective this would amount to criteria by which to determine when the specified behavior

had occurred. The description of what is done with the subject would be concerned mainly with a description of an experimental situation based on the operational definition of the specified behavior. For this reason it was important this operational definition be described as part of the problem previous to attempting to design or describe the experimental procedure.

In specifying the criteria the experimenter must take into account the previously specified level of significance and permissible level of Type II error. The method must be designed in such a way that when the criteria are satisfied that the Type II error level is not exceeded and in such a way that, if the hypothesis is rejected, then the Type I error level is not exceeded.

The final phase of the experiment involves performing the procedures defined and collecting the results.

This experiment results in one of two determinations. That is, if the criteria are satisfied then the hypothesis is accepted at the given permissible level of Type II error. The other outcome occurs when the criteria are not satisfied in which case the hypothesis is rejected at the given level of significance. Thus in measuring the achievement of one behavioral objective we have two outcomes. Either the subject is able to perform the given behavior or he is not able to perform the given behavior, each of these outcomes having their associated level of significance or level of Type II error, respectively. By structuring the evaluation of the objective as a scientific experiment the outcomes are valid and replicable.

A Generalized System of Evaluation

Measuring achievement of a collection of behavioral objectives requires a system of evaluation. A system of evaluation consists of a specified number of evaluation categories together with corresponding criteria. If an individual satisfies the criterion for a given evaluation category, then his performance would be classified in that category. For instance, given the evaluation category of A in the common grading system using A, B, C, D and E, an individual satisfying the criterion for an A would find his performance classified in that evaluation category. Another system of evaluation is the pass-fail system. This is much like the experiment described above in determining whether an individual's performance is satisfactory or unsatisfactory. That is, if the performance satisfies the criterion for the pass evaluation category, then the performance of the individual would be classified in that category. On the other hand, if the performance did not satisfy the criterion for the pass category then it would satisfy that for the fail category.

This notion of evaluation systems may be generalized by specifying an arbitrary number of evaluation categories with their corresponding criteria. Different terminal behaviors require different systems of evaluation. For example, evaluating the performance of a behavioral objective such as making a free throw in basketball would require only the pass-fail system of evaluation. In this case, the ball either goes through the hoop of the basket or it fails to go through the hoop of the basket. But, a behavior such as designing a suspension bridge between

two given points would be much more complicated. A system of evaluation suitable for evaluating the performance of this behavior would require more than two categories of evaluation because of the various aspects of the behavior involved in the design of a suspension bridge between two points. In the case of the construction of this suspension bridge, an individual's performance might be classified in any or all of the evaluation categories. That is, the evaluation categories might not be ordered in any way.

On the other hand, when a specified behavior involves a behavior chain, there are definite reasons for ordering, or ranking, the evaluation categories of the system of evaluation that is chosen to evaluate an individual's performance. For example, the behavior consisting of an actor's performance of a play with various scenes which follow naturally in sequence would best be evaluated using a system of evaluation with as many categories as there are scenes in the play. The criterion for a given category could be defined in such a way that the actor's performance is classified in the corresponding evaluation category when he had performed satisfactorily in all scenes up to a given one. These categories would increase in desirability with that category least desirable, corresponding to the first scene in the play, to that category most desirable, corresponding to the last scene in the play. The categories then would increase in desirability in the same sequence as the scenes follow one another in the play.

It might appear that it would not always be possible to order or rank the evaluation categories. However, the inferences which must be

made from the evaluations demand that the evaluation categories be at least partially ranked. This is the case where the specified terminal behavior is a complex of lesser included behaviors, some of which were performed independently. The task of constructing the suspension bridge is an example. The structural analysis that must be performed in order to determine the forces acting in the members of the bridge involves the use of numerous physical principles and mathematical or computational skills. On the other hand, the choice of materials from which to build the bridge involves a set of behaviors largely independent of those necessary in analyzing the forces in the bridge. Performing the structural analysis and determining the materials from which to build the bridge both involve lesser included behaviors which in and of themselves have value. The system of evaluation chosen to evaluate the behavior involved in designing the bridge should take into account and reflect the correct performance of these lesser included behaviors. Thus, at least some of the evaluation categories must be arranged in some order or ranked. These partial orderings can be combined in such a way as to achieve a full ordered sequence of evaluation categories by consolidating an ordered evaluation category among the partial orderings.

Since the partial orderings derive their existence from the relationship of lesser included behavior it follows that the complete ordering of the various evaluation categories reflects the ordering of the lesser included behaviors whose performance leads to the performance of the overall objective. For example if there are five evaluation categories in this final completely ordered set of evaluation categories and if

these are assigned the grades E, D, C, B and A, in ascending order, then the grades E, D, C, B and A, reflect the degree to which the individual performing the behavior is able to achieve the final terminal behavior specified. Measurements obtained from such a system of evaluation as this have more significance than similar measurements obtained in a situation where the criteria for the various evaluation categories are determined strictly on the basis of number of answers correct or other such arbitrary criterion. Additional significance accrues if the instruments embody the procedures of a valid evaluation experiment.

There are other bases for justifying an ordering or ranking of the evaluation categories. For example, Bloom's (1, 1956) taxonomy or Gagne's (2, 1970) behavioral classification system could be used to classify in this order the behaviors involved in a given course of instruction. Another rationale for ordering the evaluation categories in a given system of evaluation would be to use the "push down" principle in conjunction with Gagne's behavioral classification system. Briefly, the "push down" principle states that a behavior, once performed at a problem solving level, for instance, very well might later be performed at an analysis level or classification level according to Gagne's classification system. Thus the criterion for a given evaluation category might involve the number of Gagne's categories which the given behavior was "push down".

Implications of Systems of Evaluation

The main implication of the foregoing discussion of systems of evaluation is that performance of different behaviors is evaluated by using different systems of evaluation. Thus a unit of instruction or a short course of instruction might best be evaluated by using a pass-fail system. This would insure final competence in the performance of the terminal behavior, and would also provide a basis for teacher accountability. On the other hand, an academic program or long-range training program might better be evaluated using a system of evaluation with more categories of evaluation. Such a system of evaluation would provide means for indicating differences in levels of achievement within that particular academic training program.

Another implication of such systems of evaluation is that through their use, better bases of comparison between similar courses of instruction at different institutions can be established.

The third implication is that more meaningful feedback can be given to those individuals whose performance is being evaluated. In view of the importance placed on individual achievement in the American system of education, this third implication could very well be the most important.

Bibliography

1. Bloom, B. S. (ed.), et. al. Taxonomy of Educational Objectives Handbook I: Cognitive Domain. New York: McKay, 1956.

2. Gagné, Robert M. The Conditions of Learning. New York: Holt, Rinehart and Winston, Inc., 1970.